N° 82-003-X au catalogue
ISSN 1209-1375

Rapports sur la santé

Couplage de données de recensement et de données d'hôpital : deux approches

par Michelle Rotermann, Claudia Sanmartin, Gisèle Carrière, Richard Trudeau, Hélène St-Jean, Abdelnasser Saïdi, Alexander Reicker, Aimé Ntwari et Eric Hortop

Date de diffusion : le 15 octobre 2014

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Comment accéder à ce produit

Le produit n° 82-003-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de parcourir par « Ressource clé » > « Publications ».

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Notre organisme » > « Offrir des services aux Canadiens ».

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2014

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada (http://www.statcan.gc.ca/reference/licence-fra.htm).

This publication is also available in English.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, ses entreprises, ses administrations et les autres établissements. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

.	indisponible pour toute période de référence
..	indisponible pour une période de référence précise
...	n'ayant pas lieu de figurer
0	zéro absolu ou valeur arrondie à zéro
0[s]	valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
[p]	provisoire
[r]	révisé
x	confidentiel en vertu des dispositions de la *Loi sur la statistique*
[E]	à utiliser avec prudence
F	trop peu fiable pour être publié
*	valeur significativement différente de l'estimation pour la catégorie de référence (p<0,05)

Couplage de données de recensement et de données d'hôpital : deux approches

par Michelle Rotermann, Claudia Sanmartin, Gisèle Carrière, Richard Trudeau, Hélène St-Jean, Abdelnasser Saïdi, Alexander Reicker, Aimé Ntwari et Eric Hortop

Résumé

Contexte

La présente étude compare deux approches pour le couplage de données, l'une fondée sur un registre et l'autre non fondée sur un registre. Des données pour le Manitoba et l'Ontario provenant du Recensement de la population de 2006 ont été couplées aux données sur l'hospitalisation tirées de la Base de données sur les congés des patients (BDCP).

Données et méthodes

À partir d'un couplage probabiliste, on a utilisé l'approche *fondée sur un registre* pour le couplage des données du recensement à celles des registres provinciaux de l'assurance-maladie, suivi par un couplage déterministe avec la BDCP à partir du numéro d'assurance-maladie (NAM). L'approche *non fondée sur un registre* repose sur un appariement exact hiérarchique déterministe à partir de trois variables communes aux deux fichiers pour le couplage des données du recensement et de celles de la BDCP. Les approches ont été comparées du point de vue des taux de couplage et de couverture, de la sensibilité et de la spécificité, ainsi que de l'uniformité des NAM dans les enregistrements couplés.

Résultats

Les résultats des couplages selon l'approche fondée sur un registre et selon l'approche non fondée sur un registre sont similaires. Au Manitoba, 7 % et 6 % des répondants au questionnaire détaillé du recensement ont été couplés à la BDCP suivant l'approche fondée sur un registre et l'approche non fondée sur un registre, respectivement; en Ontario, le taux de couplage a été de 5 % pour les deux approches. Selon l'approche fondée sur un registre, les données du recensement et de la BDCP couplées représentaient 84 % (données pondérées) des admissions à l'hôpital enregistrées dans la BDCP de 2006-2007 en Ontario et au Manitoba, comparativement à 82 % au Manitoba et en Ontario selon l'approche non fondée sur un registre.

Interprétation

Faute de pouvoir accéder aux registres provinciaux de l'assurance-maladie, avec lesquels peuvent être couplées les données du recensement, on peut avoir recours à une approche non fondée sur un registre pour créer un ensemble de données dont la qualité est appropriée pour la recherche.

Mots-clés

Couverture, collecte de données, couplage de données, bases de données, dossiers médicaux, dossiers d'hôpital, analyse de métadonnées.

Auteurs

Michelle Rotermann (michelle.rotermann@statcan.gc.ca), Claudia Sanmartin et Gisèle Carrière travaillent à la Division de l'analyse de la santé, Richard Trudeau, à la Division de la statistique de la santé, et Hélène St-Jean, Abdelnasser Saïdi, Alexander Reicker, Aimé Ntwari et Eric Hortop, à la Division des méthodes d'enquêtes auprès des ménages de Statistique Canada, Ottawa (Ontario) K1A 0T6.

Le couplage d'enregistrements, c'est-à-dire le processus d'appariement d'enregistrements — entre des ensembles de données différents ou à l'intérieur d'un seul —, est courant dans la recherche sur la santé[1-7]. Il sert à créer un ensemble de données enrichies dont les applications sont plus larges[6-11]. Les données qui s'y prêtent sont les données complémentaires, c'est-à-dire se trouvant dans une source de données sur deux.

Pour produire des résultats exacts, le couplage de données doit se baser sur un identificateur unique commun[12,13]. Un identificateur doit s'appliquer à une seule personne, et une personne doit posséder un seul identificateur[14]. Dans les couplages du domaine de la santé au Canada, on a eu recours à ce titre au numéro d'assurance-maladie provinciale (NAM)[2,3,11,15]. Ce numéro n'est toutefois pas employé dans la plupart des bases de données (p. ex., statistique de l'état civil, recensement et dossiers de l'impôt).

À défaut de certaines données, on peut recourir à une approche *fondée sur un registre* pour le couplage de données. Par exemple, les registres de l'assurance-maladie peuvent servir de fichiers de « rapprochement » pour le couplage, parce qu'on y retrouve les NAM, ainsi que les noms et d'autres variables d'identification[2,3,11,15]. Mais ces registres ne sont pas toujours accessibles. On peut également adopter une *approche non fondée sur un registre*, laquelle repose sur l'appariement d'enregistrements provenant de bases de données différentes, à partir de combinaisons de données personnelles co-occurrentes, comme la date de naissance et le code postal[13,14,16,17].

Lorsque les bases de données ont trait à la même population, on s'attend généralement à ce que la plupart des enregistrements puissent être couplés. Toutefois, lorsqu'on prévoit qu'une fraction seulement des enregistrements seront couplés, la prédétermination d'un taux de couplage raisonnable pose un problème. Cela est typique dans les couplages axés sur la santé — par exemple, un petit nombre de personnes seront hospitalisées ou décéderont pendant le suivi. Dans de tels cas, les couplages ont été évalués en comparant les résultats d'approches différentes à partir des mêmes jeux de données[12,14,18-20]; en comparant les taux et les répartitions en pourcentage à l'égard de variables pour lesquelles existent des données couplées et non couplées[14,21,22]; et en calculant la sensibilité et la spécificité[19,23,24]. Selon des études canadiennes, pour des enregistrements

qui, selon toute vraisemblance, seront couplés, un taux d'appariement d'environ 75 % est considéré acceptable pour la recherche[5,25,26].

La présente étude compare deux approches (l'une *fondée sur un registre* et l'autre non) pour le couplage de données du Recensement de la population de 2006 et de données d'hôpital tirées de la Base de données sur les congés des patients (BDCP) pour le Manitoba et l'Ontario, deux provinces pour lesquelles Statistique Canada a accès aux données des registres de l'assurance-maladie (NAM). Elle a pour but de déterminer s'il est possible de produire un ensemble de données dont le niveau de qualité est assez élevé pour la recherche, sans se servir de données de « rapprochement » comme celles provenant des registres provinciaux de l'assurance-maladie. Le couplage a été approuvé par le Comité des politiques[27] et régi par la Politique sur le couplage d'enregistrements de Statistique Canada[28].

Données et méthodes

Sources des données

Recensement de la population de 2006

Les données du Recensement de 2006 ont été recueillies à l'aide d'un questionnaire abrégé (court) et d'un questionnaire détaillé (long). La population au complet a répondu aux sept questions de base que comprenait le questionnaire abrégé, portant sur des caractéristiques démographiques comme la date de naissance, le sexe et le code postal de chaque membre du ménage[29]. Environ 20 % des ménages privés ont en outre été sélectionnés au hasard pour répondre au questionnaire détaillé, lequel comportait 52 questions additionnelles sur le revenu, la scolarité, l'origine ethnique, et le statut d'Autochtone, entre autres[29].

Le fichier de données sur les répondants au questionnaire abrégé du recensement renferme des données sur les répondants au questionnaire détaillé, et comporte donc des enregistrements pour la presque totalité de la population (97 % de celle du Manitoba et 96 % de celle de l'Ontario). Le fichier de données sur les répondants au questionnaire détaillé ne porte que sur les ménages sélectionnés pour répondre au questionnaire détaillé (20 %).

Avant que n'ait lieu le couplage, le fichier du recensement a été épuré, nettoyé de ses enregistrements en double et validé[29,30]. Le fichier correspondant au questionnaire abrégé a été soit couplé aux registres provinciaux de l'assurance-maladie, puis à la BDCP (approche *fondée sur un registre*), soit couplé directement à la BDCP (approche *non fondée sur un registre*). L'inclusion dans le couplage de tous les enregistrements du recensement a permis de repérer les enregistrements provenant des registres provinciaux de l'assurance-maladie et de la BDCP pour les répondants qui avaient rempli le questionnaire abrégé uniquement et de supprimer ces enregistrements. Ont été gardées dans la cohorte de l'étude seuls les enregistrements pour les personnes qui avaient répondu au questionnaire détaillé du recensement.

Registres provinciaux de l'assurance-maladie (approche fondée sur un registre)

Dans l'approche *fondée sur un registre*, on a utilisé le fichier d'enregistrements du Régime d'assurance-maladie du Manitoba (RAMM) et la Base de données des personnes inscrites (BDPI) de l'Ontario comme fichiers de « rapprochement ». Une fois les données du recensement couplées avec celles d'un registre provincial de l'assurance-maladie, il devient possible de coupler les données du recensement avec celles de la BDCP, grâce à la concordance des NAM.

Le RAMM et la BDPI comprennent les enregistrements des personnes éligibles pour recevoir des services de santé, au Manitoba et en Ontario respectivement. Étant donné que l'inscription n'exige aucuns frais, la couverture de la population est élevée[31-33]. Les enregistrements pour les personnes qui ne vivent plus dans la province, mais qui demeurent couvertes jusqu'à trois mois après leur déménagement, sont inclus dans les deux registres. Chez les nouveaux résidents, il y a une période d'attente de trois mois pour être couvert. Les personnes couvertes par un autre régime (par exemple, les détenus, ainsi que les membres de la GRC et des Forces canadiennes) sont exclues.

Avant d'être couplées avec les données du recensement, les données du RAMM et de la BDPI ont fait l'objet d'un prétraitement, lors duquel on a repéré les cas de personnes ayant plusieurs NAM (Manitoba = 0,2 % ou 3 588 cas; Ontario = 1 % ou 165 123 cas). Seuls étaient conservés dans la cohorte de l'étude les enregistrements comportant un nom (de famille), une date de naissance antérieure au 1er janvier 2007, une assurance valide au cours de la période du 31 décembre 2005 au 1er janvier 2007 et, le cas échéant, une date de décès postérieure au 31 décembre 2005[36].

Base de données sur les congés des patients de 2006-2007

La BDCP comprend des données démographiques, administratives (y compris le NAM) et cliniques pour tous les congés reçus des hôpitaux de soins de courte durée, ainsi que de certains hôpitaux de soins psychiatriques, de soins de réadaptation pour malades chroniques et de soins de longue durée, ainsi que de certains services de chirurgie d'un jour[34-35]. La version de 2006-2007 de la BDCP porte sur les congés d'hôpital survenus du 1er avril 2006 au 31 mars 2007 inclusivement (n = 3 186 079).

Fichiers maîtres des particuliers T1 pour 2005-2006-2007 (approche non fondée sur un registre)

Afin de tenir compte des changements de code postal survenus avec le temps et d'améliorer la qualité du couplage avec la BDCP selon l'approche *non fondée sur un registre*, les codes postaux tirés des dossiers de l'impôt de Statistique Canada pour 2005, 2006 et 2007 (Fichiers maîtres des particuliers T1 – FMPT1) ont été versés au fichier de données sur les personnes ayant rempli le questionnaire abrégé du recensement. À partir du sexe, de la date de naissance et de noms et prénoms partiels, la plupart des enregistrements du recensement (91 %) ont pu être couplés à au moins une année de données sur l'impôt. Dans le

cas des personnes qui ne produisent pas de déclaration de revenu chaque année et (ou) qui ne sont pas tenues de le faire (p. ex. les enfants), les codes postaux ont été repérés et attribués grâce aux renseignements fournis par les autres membres déclarants du ménage.

Couplage d'enregistrements

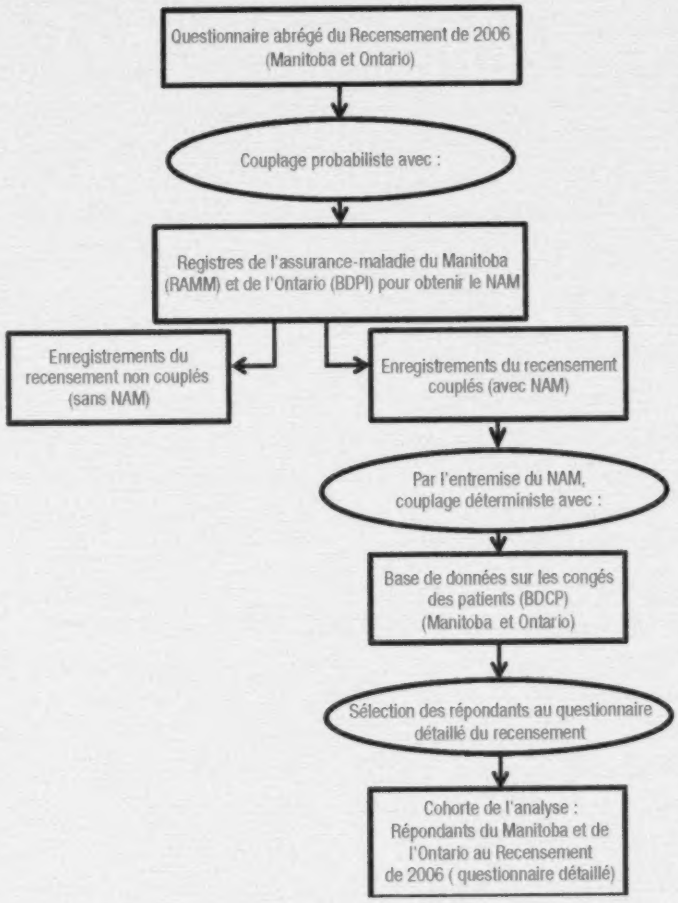
Approche fondée sur un registre

Le couplage *fondé sur un registre* a été effectué en deux étapes. Premièrement, les enregistrements du recensement tirés du questionnaire abrégé pour le Manitoba et l'Ontario ont été couplés par une méthode probabiliste aux registres provinciaux de l'assurance-maladie, en vue d'obtenir les NAM. Deuxièmement, à partir des NAM, les enregistrements du recensement tirés du questionnaire détaillé appariés à ces registres ont été couplés par une méthode déterministe à la BDCP (figure 1).

Des scores de probabilité fondés sur les similarités quant à la date de naissance, au code postal, au sexe, au nom et au prénom ont servi à estimer la probabilité que les enregistrements appariés représentent la même personne[19,37,38]. Des poids (positifs/négatifs) ont été attribués aux champs de comparaison et leur somme a servi à créer un poids de couplage global. Des seuils distincts servant à distinguer les appariements vrais des non-appariements ont été prédéterminés pour le Manitoba et l'Ontario, à partir des répartitions de ces poids. En raison de la taille des fichiers de l'Ontario, le fait de comparer chaque enregistrement du recensement avec chaque enregistrement du registre était prohibitif. Par conséquent, on a fractionné le fichier de l'Ontario, de sorte que seuls les enregistrements pour un même sexe ont été comparés. On n'a pas procédé ainsi pour le Manitoba parce que la population et les fichiers étaient de plus petite taille.

On a classé les paires d'enregistrements d'après leurs scores. On a considéré que celles dont le score était supérieur au seuil prédéterminé constituaient un appariement. Avant de fixer définitivement les seuils, on a examiné les paires d'enregistrements s'en rapprochant et les seuils ont été corrigés au besoin.

Figure 1
Couplage fondé sur un registre, Manitoba et Ontario



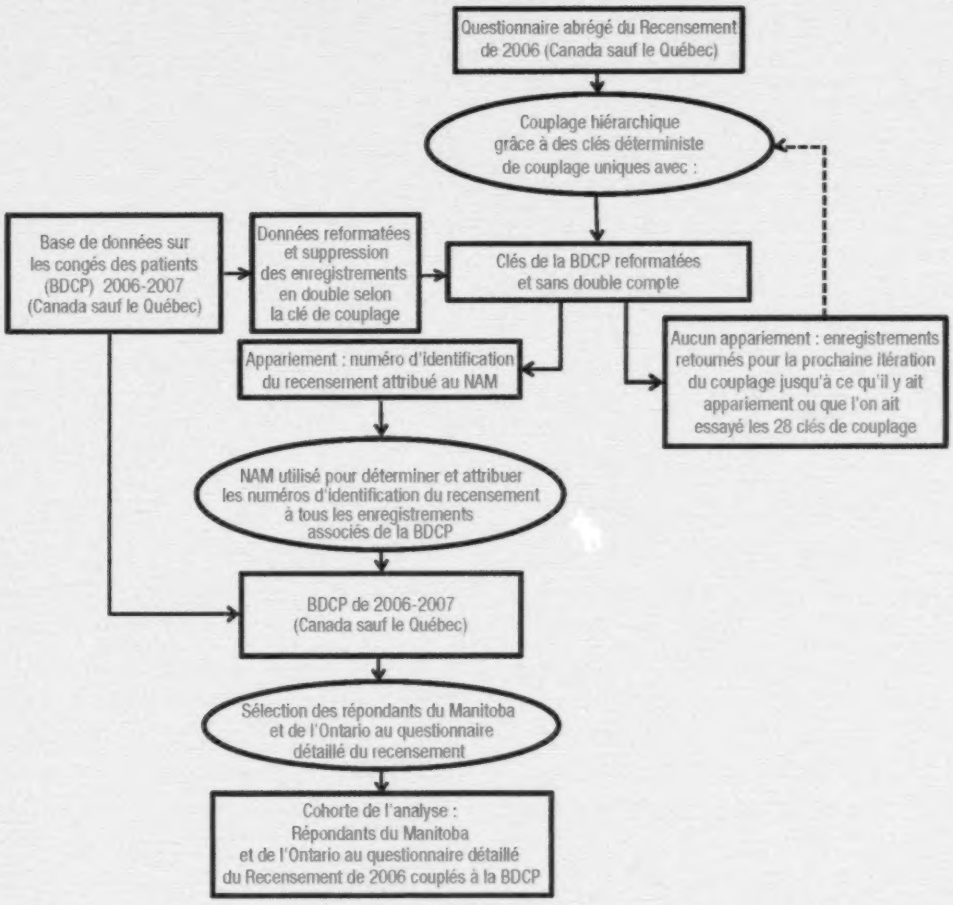
Les fichiers recensement–registres couplés renfermant les identificateurs du recensement ainsi que les NAM ont ensuite servi au couplage par une méthode déterministe des identificateurs du recensement aux fichiers de la BDCP. Seuls les identificateurs du recensement correspondant aux personnes ayant rempli le questionnaire détaillé ont été conservés pour l'analyse[38].

Approche non fondée sur un registre

L'approche *non fondée sur un registre* avait recours à l'appariement exact hiérarchique déterministe de clés de couplage constituées de combinaisons de trois variables communes au questionnaire abrégé du recensement et aux enregistrements de la BDCP (date de naissance, code postal et sexe) (figure 2). L'appariement a nécessité la comparaison de paires d'enregistrements recensement–BDCP, afin de déterminer si elles avaient trait à la même personne. Si seuls les appariements utilisant une clé avaient été acceptés, les taux de couplage auraient été plus faibles[12]. En utilisant plusieurs clés successivement, l'appariement exact hiérarchique déterministe représente une précision qui permet de maximiser la puissance de discrimination des données de couplage et de réduire l'incidence des valeurs manquantes et des erreurs[16].

Les données ont été reformatées en vue du couplage. Lorsque les enregistrements n'indiquaient pas le sexe (n ~ 312 000, recensement; n ~ 300, BDCP), on a attribué un sexe à l'enregistrement existant et créé un autre enregistrement comportant la même date de naissance et le même code postal, mais pour le sexe opposé. On a également créé un autre enregistrement lorsque le code postal tiré du recensement était différent de celui issu du traitement (n ~ 525 000).

Figure 2
Couplage non fondé sur un registre, Canada sauf le Québec



On a repéré les enregistrements multiples pour une seule personne dans les fichiers du recensement et de la BDCP, à partir du numéro d'identification de groupe et de la clé NAM (NAM + province), ce qui a facilité leur suppression une fois le couplage terminé.

Selon le couplage *non fondé sur un registre*, on a utilisé une approche itérative selon laquelle on a appliqué 28 règles aux fichiers recensement–BDCP, une à la suite de l'autre. Les règles dans les premières itérations étaient strictes; les subséquentes admettaient des divergences. Par exemple, la première itération exigeait un appariement exact entre les enregistrements du recensement et de la BDCP pour la date de naissance, le sexe et le code postal. Dans les itérations 2 à 4, un appariement exact était nécessaire pour la date de naissance, le sexe et le code postal entre la BDCP et les FMPT1 pour 2005, 2006 et 2007. Aux itérations 5 à 10, les règles visant le code postal s'assouplissaient, ce qui permettait de laisser tomber l'un des six caractères du code. Ce processus a été répété en utilisant les codes postaux provenant des FMPT1 pour 2005, 2006 et 2007 (itérations 11 à 28). Après chaque itération, les enregistrements du recensement comportant le même numéro d'identification de groupe et ceux de la BDCP indiquant la même clé NAM que dans le couplage ont été supprimés d'itérations subséquentes, afin que les sujets pris en compte dans les fichiers du recensement et dans les fichiers de la BDCP à cause de clés de couplage multiples, ne soient couplées qu'une fois.

Il était possible, dans chaque ensemble de données, d'avoir des enregistrements comportant des clés de couplage en double, en particulier dans la BDCP, car pour une année donnée, une personne pouvait avoir plusieurs dossiers d'hôpital portant la même date de naissance, le même sexe et le même code postal. Afin de gagner en efficacité et d'éliminer tout résultat d'égalité éventuel, avant d'entreprendre le couplage on a supprimé les enregistrements en double des fichiers du recensement et de la BDCP à l'aide de cette clé de couplage. Une fois le couplage terminé, les enregistrements du recensement qui avaient été ajoutés pour tenir compte des codes postaux du recensement incohérents et (ou) du sexe manquant ont été supprimés du fichier, et les enregistrements de la BDCP que l'on avait laissés de côté (pour hospitalisations

multiples de la même personne) y ont été rajoutés. Enfin, seuls les enregistrements associés à l'hospitalisation de résidents du Manitoba et de l'Ontario ont été conservés aux fins de la présente étude comparative.

Protéger la confidentialité des données du répondant

Statistique Canada s'assure de protéger la confidentialité des données du répondant au cours du processus de couplage et de l'utilisation subséquente des fichiers de données couplées. Seuls les employés impliqués directement dans le processus de couplage ont accès aux identificateurs uniques requis pour le couplage (par exemple les noms et le numéro d'assurance-maladie) et ils n'ont pas accès à l'information qui à trait à la santé. Lorsque le processus de couplage est terminé, un fichier analytique de données est créé dans lequel tous les identificateurs personnels sont supprimés. L'accès à ce fichier libre d'identificateurs est réservé aux analystes aux fins de validation et d'analyses.

Nombres d'enregistrements

Pour l'approche *fondée sur un registre*, le nombre d'enregistrements tirés du questionnaire abrégé du recensement qui ont été utilisés pour le couplage était de 1 111 133 pour le Manitoba et de 11 704 729 pour l'Ontario. Du RAMM, on a pris 1 201 152 enregistrements et de la BDPI, 13 121 593 enregistrements. De ceux-ci, 246 578 des 278 937 enregistrements tirés du questionnaire détaillé pour le Manitoba et 2 136 455 des 2 387 911 enregistrements qui en ont été tirés pour l'Ontario ont été couplés à un fichier de l'assurance-maladie. Ceux-ci étaient donc admissibles au couplage avec la BDCP *fondé sur un registre* (tableau 1).

Étant donné que l'on n'avait pas besoin des registres provinciaux de l'assurance-maladie pour l'approche *non fondée sur un registre*, la plupart des enregistrements tirés du questionnaire abrégé du recensement pour les résidents de la plupart des provinces/territoires étaient admissibles au couplage avec la BDCP (n = 23 592 671). Les enregistrements du recensement ont été exclus s'ils renfermaient une date de naissance ou un code postal non valide ou manquant, ou encore si le code postal au moment du recensement et le code postal après le recensement se rapportaient à une adresse au Québec. (Les personnes du Québec dénombrées au recensement n'étaient pas admissibles au couplage parce que Statistique Canada n'a pas accès aux dossiers sur les hospitalisations au Québec.) La cohorte définitive pour l'étude *non fondée sur un registre* comptait 278 937 enregistrements tirés du questionnaire détaillé du recensement pour le Manitoba et 2 387 911 enregistrements pour l'Ontario.

Selon l'une ou l'autre approche, les enregistrements de la BDCP comportant une date de naissance après le 16 mai 2006 (jour du recensement) étaient exclus, tout comme les enregistrements se rapportant aux personnes résidant à l'étranger et aux mortinaissances. Aux fins du couplage *fondé sur un registre*, seuls les enregistrements de la BDCP comportant un NAM valide pour le Manitoba (227 069) ou l'Ontario (1 081 443) ont été retenus. En ce qui concerne le couplage *non fondé sur un registre*, on n'a retenu que les enregistrements de la BDCP pour lesquels il ne manquait aucun renseignement sur la date de naissance et le code postal (2 106 104).

Validation

Taux de couplage

Pour évaluer les taux de couplage obtenus selon l'approche avec *registre*, on a examiné les pourcentages de répondants au recensement qui ont été couplés au RAMM ou à la BDPI. Compte tenu des différences de couverture au recensement et dans les fichiers des registres provinciaux de l'assurance-maladie, les taux de couplage devraient être voisins de 100 %, sans toutefois l'atteindre.

Selon les deux approches (*avec* et *sans registre*), on a calculé les taux de couplage (global et selon certaines caractéristiques sociodémographiques) avec la BDCP pour les répondants ayant rempli le questionnaire détaillé du recensement au Manitoba et en Ontario. Ces taux devraient refléter la prévalence d'au moins une hospitalisation au cours de l'exercice 2006-2007, et on s'attend à ce qu'ils soient plus élevés chez les groupes de personnes dont la probabilité d'être hospitalisé est plus élevée (p. ex. les personnes âgées).

Exactitude du couplage

On a calculé la sensibilité (vrais positifs) et la spécificité (vrais négatifs) pour évaluer l'exactitude du couplage au niveau de l'enregistrement. Les résultats fondés sur l'approche du *registre* ont été utilisés à titre de « norme de référence » pour la comparaison des résultats de l'approche *sans registre*[2,3,11,15].

Dans le cas des enregistrements du recensement pour lesquels les deux approches ont donné lieu à un couplage à la BDCP, on a comparé les NAM pour évaluer la cohérence interne. Les appariements constituaient la preuve de couplages exacts.

Analyse de la couverture

On a calculé les taux de couverture pour chaque approche en divisant le nombre de congés d'un hôpital de soins de courte durée qu'avaient reçus les répondants au questionnaire détaillé du recensement au Manitoba et en Ontario d'après les données du couplage recensement–BDCP (numérateur) par le nombre de congés de ce genre déclarés dans les données non couplées de la BDCP pour 2006-2007 (dénominateur). On a calculé les taux de couverture non pondérés et pondérés pour chaque approche, y compris les taux globaux et ceux se rapportant aux hospitalisations attribuables à trois diagnostics « principaux » : maladies du système circulatoire, lésions traumatiques et empoisonnements, ainsi que grossesse, accouchement et puerpéralité.

Les inférences au niveau de la population doivent être basés sur des poids[39]. Après l'application des poids, les taux de couverture devraient se rapprocher de 100 %, sans toutefois l'atteindre. Sans l'application de poids, les taux de couverture devraient se situer autour du pourcentage de la population ayant répondu au questionnaire détaillé du recensement (25 % au Manitoba; 20 % en Ontario). Compte tenu des différences de couverture de la population dans les

Tableau 1
Nombre et pourcentage de répondants au Recensement de 2006 dont les enregistrements ont été couplés à ceux du Régime d'assurance-maladie du Manitoba (RAMM) ou de la Base de données des personnes inscrites (BDPI) de l'Ontario, selon certaines caractéristiques sociodémographiques, Manitoba et Ontario, 2006-2007

	Répondants au recensement (Manitoba)			Répondants au recensement (Ontario)		
	Total	Couplage avec le RAMM		Total	Couplage avec la BDPI	
Caractéristiques	'000	'000	%	'000	'000	%
Questionnaire abrégé, total	**1 111,1**	**1 038,9**	**93**	**11 704,7**	**10 578,0**	**90**
Questionnaire détaillé, total	**278,9**	**246,6**	**88**	**2 387,9**	**2 136,5**	**89**
Sexe						
Hommes	138,2	120,9	87	1 169,0	1 043,5	89
Femmes	140,8	125,7	89	1 219,0	1 092,9	90
Groupe d'âge						
Moins d'un an	4,1	3,5	85	26,1	23,9	91
1 à 4 ans	15,9	13,6	86	107,3	97,9	91
5 à 9 ans	21,1	18,3	87	143,4	131,1	91
10 à 14 ans	23,1	20,2	87	163,2	148,2	91
15 à 19 ans	22,2	19,4	87	166,6	147,9	89
20 à 24 ans	19,3	16,2	84	155,4	133,7	86
25 à 34 ans	34,6	30,2	87	299,4	261,3	87
35 à 44 ans	38,8	34,6	89	377,5	335,6	89
45 à 54 ans	39,2	35,4	90	369,4	331,5	90
55 à 64 ans	28,2	25,7	91	269,9	245,3	91
65 à 74 ans	17,1	15,6	92	171,2	155,5	91
75 ans et plus	15,4	13,8	90	138,3	124,5	90
Quintile de revenu de la famille économique						
1 (le plus faible)	55,8	45,5	82	477,4	401,8	84
2	55,8	48,2	86	477,7	425,5	89
3	55,8	49,9	89	477,6	432,7	91
4	55,8	51,3	92	477,6	436,6	91
5 (le plus élevé)	55,8	51,6	93	477,6	439,8	92
Scolarité (18 ans et plus)						
Diplôme d'études secondaires ou niveau supérieur	135,6	122,8	91	1 494,9	1 338,0	90
Pas de diplôme d'études secondaires	65,1	56,0	86	350,9	306,2	87
Situation d'emploi						
Occupé(e)	127,1	114,7	90	1 216,2	1 089,9	90
En chômage	10,4	8,9	85	84,9	74,3	87
Inactif(ve)	141,4	123,0	87	1 086,8	972,3	89
Connaissance des langues officielles						
Anglais seulement	254,2	224,5	88	2 053,7	1 841,3	90
Français seulement	0,4	0,4	89	9,5	8,5	90
Anglais et français	20,9	19,0	91	271,7	242,7	89
Ni l'anglais ni le français	3,4	2,7	81	53,0	44,0	83
Identité autochtone						
Toute identité autochtone	81,7	68,0	83	86,6	71,7	83
Indien de l'Amérique du Nord seulement	65,2	53,3	82	69,2	56,4	81
Métis seulement	15,7	14,0	89	15,1	13,4	89
Inuit seulement	0,1	0,1	76	0,4	0,3	83
Identités autochtones multiples ou autres	0,7	0,6	83	1,9	1,6	86
Non-Autochtone	197,3	178,5	91	2 301,3	2 064,7	90
Statut d'immigrant						
Immigrant	29,5	26,3	89	657,3	578,5	88
Non-immigrant	247,9	219,5	89	1 704,1	1 545,6	91
Résident non permanent	1,6	0,8	47	26,5	12,4	47
Mobilité résidentielle sur un an						
Même adresse	241,2	215,1	89	2 044,3	1 846,6	90
Adresse différente au Canada	31,5	26,5	84	286,8	246,3	86
Déménagé depuis l'extérieur du Canada	2,0	1,5	75	26,9	19,6	73
Résidence en région rurale/urbaine						
Rurale et agricole	120,5	103,5	86	395,1	353,7	90
Petit centre de population (1 000 à 29 999 habitants)	26,0	23,7	91	218,0	199,2	91
Centre de population moyen (30 000 à 99 999 habitants)	7,8	6,9	89	186,0	169,2	91
Grand centre de population (100 000 habitants et plus)	124,7	112,4	90	1 588,7	1 414,3	89

Taux de couplage = (Répondants au recensement couplés au RAMM ou à la BDPI/Nombre total de répondants au recensement)*100
Notes : Estimations de la BDPI et du RAMM couplées au recensement sur la base des répondants au questionnaire détaillé du recensement seulement. Sont exclus les enregistrements concernant les personnes vivant en établissement et les enregistrements n'ayant pas été couplés au RAMM ou à la BDPI.
Sources : Recensement de la population de 2006; Base de données des personnes inscrites et Régime d'assurance-maladie du Manitoba pour 2006-2007 couplés au recensement.

données recensement–BDCP couplées et les données de la BDCP non couplées, on ne s'attend pas à une correspondance exacte. Par exemple, on tient compte des personnes vivant en établissement – de gros utilisateurs de services hospitaliers[10,41] – dans les données non couplées de la BDCP, mais pas dans les données résultant du couplage recensement–BDCP. Les poids du recensement n'ont pas été corrigés de ces irrégularités, si bien que leur application pourrait fausser certaines estimations.

Afin que la population de la BDCP ressemble davantage à la population cible des fichiers couplés, on a soustrait du dénominateur les enregistrements de la BDCP pour les populations suivantes : les résidents de maisons pour personnes âgées (2 368 congés pour le Manitoba et 24 487 pour l'Ontario, selon la BDCP), les personnes nées après le jour du recensement, les mortinaissances et les non-Canadiens.

Caractéristiques sociodémographiques

Pour chaque famille économique ou personne seule, on a déterminé le revenu total après impôt provenant de toutes les sources et pour tous les membres de la famille, que l'on a corrigé pour tenir compte de la composition et de la taille de la famille et séparé en *quintiles de revenu*.

Le *niveau le plus élevé de scolarité* des personnes de 18 ans et plus a été défini en fonction de deux catégories : diplôme d'études secondaires ou pas de diplôme d'études secondaires.

La *situation d'emploi* a été définie en fonction des catégories « occupé(e) », « en chômage » ou « inactif(ve) ».

Les répondants ont été divisés en quatre catégories, selon leur connaissance autodéclarée des *langues officielles* du Canada : français seulement, anglais seulement, français et anglais, ni le français ni l'anglais.

Les données sur le *statut d'Autochtone* ont été tirées de la question : « Cette personne est-elle un Autochtone, c'est-à-dire un Indien de l'Amérique du Nord, un Métis ou un Inuit (Esquimau)? » Les répondants devaient indiquer toutes les réponses qui s'appliquaient. Les réponses ont été groupées en six catégories : Indien de l'Amérique du Nord (seulement), Métis (seulement), Inuit (seulement), autre Autochtone (origines multiples ou indéterminées), Autochtone (total des quatre catégories précédentes), et non-Autochtone.

Le pays de naissance, la citoyenneté et la situation à l'égard de l'immigration ont été groupés en une variable du *statut d'immigrant*, dont les catégories étaient : immigrant, non immigrant ou résident non permanent.

On a créé une variable de la *mobilité résidentielle sur un an* pour refléter les changements d'adresse. Elle comprenait les catégories « même adresse », « déménagement au Canada » ou « déménagement depuis l'extérieur du Canada ». On l'a calculée en comparant pour chaque répondant la municipalité et la province de résidence le jour du recensement et un an plus tôt.

Une variable de la résidence en région *rurale/urbaine* a permis de refléter le lieu de résidence et la taille de la collectivité. Les régions de résidence agricoles ou non agricoles comptant une population de moins de 1 000 habitants ont été considérées comme rurales/agricoles. Les autres régions de résidence comprenaient les petits centres de population (de 1 000 à 29 999 habitants), les centre de population moyens (de 30 000 à 99 999 habitants) et les gros centres de population (100 000 habitants et plus).

Résultats : hospitalisations

D'après la BDCP, on a déterminé le nombre total d'*hospitalisations toutes causes confondues* pour lesquelles les dates du congé allaient du 1er avril 2006 au 31 mars 2007 inclusivement, pour le Manitoba et l'Ontario.

Les hospitalisations attribuables à une *maladie du système circulatoire* devaient comporter un diagnostic principal (DxP) dont le code de la *Classification internationale des maladies, 10e* révision, se situait entre J00 et J99. Les hospitalisations imputables *aux lésions traumatiques et empoisonnements* étaient celles correspondant aux codes S00-S99 ou T00-T98. Les hospitalisations attribuables à la *grossesse, l'accouchement et la puerpéralité* correspondaient aux codes O00-O99.

Résultats du couplage

Recensement–registres provinciaux de l'assurance-maladie (approche fondée sur un registre)

Dans le cas de l'approche *fondée sur un registre*, 88 % (246 578) des répondants du Manitoba au recensement (questionnaire détaillé) et 89 % (2 136 455) des répondants de l'Ontario (questionnaire détaillé) ont été couplés au RAMM et à la BDPI, respectivement, et ont ainsi constitué la cohorte de l'étude admissible pour le couplage à la BDCP à partir d'un registre (tableau 1). (Les pourcentages de répondants ayant rempli le questionnaire abrégé du recensement qui ont été couplés aux registres étaient quelque peu plus élevés, soit 93 % pour le Manitoba et 90 % pour l'Ontario.)

Les taux de couplage des enregistrements des répondants au questionnaire détaillé du recensement à ceux des registres de l'assurance-maladie variaient selon les caractéristiques sociodémographiques. Par exemple, les taux au Manitoba allaient de 85 % pour les enfants de moins d'un an à 92 % pour les personnes de 65 à 74 ans. Les taux de couplage étaient relativement faibles pour les personnes du quintile de revenu inférieur, les personnes ayant déclaré ne connaître ni l'une ni l'autre des langues officielles du Canada, les résidents non permanents et les personnes qui ne vivaient pas au Canada l'année précédant le recensement. Les taux de couplage chez les groupes autochtones allaient de 76 % à 89 %.

Recensement–BDCP (approche non fondée sur un registre)

Dans le cas de l'approche *non fondée sur un registre*, 80 % ou 1,69 million de clés de la BDCP (Canada, sauf le Québec) ont été couplées à un enregistrement tiré du questionnaire abrégé du recensement (tableau 2). La majorité des couplages entre le recensement et la BDCP ont été

obtenus à la première itération (72 % ou 1,52 million), laquelle exigeait un appariement exact pour la date de naissance, le sexe et le code postal. À partir des données sur le code postal tirées des dossiers de l'impôt FMPT1, les itérations 2 à 4 ont donné lieu à 80 000 couplages de plus (4 %), et les itérations 5 à 28, à 68 000 couplages additionnels (3 %).

Comparaison des approches

Les pourcentages de répondants du Manitoba et de l'Ontario au questionnaire détaillé du recensement pour lesquels un couplage à la BDCP a eu lieu étaient similaires pour l'une et l'autre approche (tableau 3). Selon l'approche *fondée sur un registre*, 7 % des Manitobains avaient été hospitalisés dans un établissement de soins de courte durée; selon l'approche *non fondée sur un registre*, le pourcentage était de 6 %. En Ontario, le taux pour les deux approches était de 5 %.

Comme on pouvait s'y attendre, les données couplées reflétaient des différences d'utilisation des services hospitaliers. Par exemple, peu importe la province ou l'approche utilisée pour le couplage, un pourcentage plus élevé de femmes que d'hommes avaient été hospitalisées. Les enfants de moins d'un an et les personnes âgées étaient plus susceptibles que les personnes d'autres groupes d'âge d'être couplés aux dossiers d'hôpital. D'autres caractéristiques corrélées à l'âge et (ou) à l'incapacité, comme le fait d'être inactif, étaient associées à des taux d'hospitalisation plus élevés.

Sensibilité et spécificité

Les résultats du couplage avec la BDCP étaient les mêmes pour la majorité des enregistrements tirés du questionnaire détaillé du recensement, peu importe l'approche. La sensibilité et la spécificité des couplages d'enregistrements pour le Manitoba étaient de 87,9 % et 98,8 % respectivement; les chiffres correspondants pour l'Ontario étaient 89,4 % et 99,6 % (tableau 4).

L'uniformité des NAM parmi les répondants au recensement ayant été couplés à la BDCP selon l'une ou l'autre approche était également élevée. À peu près tous (99 %) les répondants du Manitoba (24 487) et de l'Ontario (106 968) au questionnaire détaillé du recensement qui ont été couplés à la BDCP selon les deux approches ont été liés au même NAM selon chaque approche.

Tableau 2
Clés ayant servi à l'appariement exact hiérarchique déterministe des enregistrements de la Base de données sur les congés des patients (BDCP) de 2006-2007 et de ceux du Recensement de 2006, et pourcentage et nombre cumulés d'enregistrements de la BDCP couplés, selon l'itération, Canada sauf le Québec

		Clés de la BDCP couplées	
Description de la clé		% cumulé	Nombre cumulé
Itération	**Codes postaux utilisés, complets**		
1	date de naissance ll sexe ll code postal RTA UDL	72,2	1 521 386
2	date de naissance ll sexe ll code postal FMPT1 RTA UDL de 2005	74,7	1 572 846
3	date de naissance ll sexe ll code postal FMPT1 RTA UDL de 2006	75,9	1 597 970
4	date de naissance ll sexe ll code postal FMPT1 RTA UDL de 2007	76,1	1 601 710
	Codes postaux utilisés, incomplets		
5	date de naissance ll sexe ll code postal du recensement RTA xDL	76,9	1 619 231
6	date de naissance ll sexe ll code postal du recensement RTA UxL	78,3	1 649 492
7	date de naissance ll sexe ll code postal du recensement RTA UDx	78,6	1 655 680
8	date de naissance ll sexe ll code postal du recensement xTA UDL	79,3	1 670 817
9	date de naissance ll sexe ll code postal du recensement RxA UDL	79,4	1 672 933
10	date de naissance ll sexe ll code postal du recensement RTx UDL	79,6	1 676 645
11	date de naissance ll sexe ll code postal FMPT1 RTA xDL de 2005	79,7	1 677 620
12	date de naissance ll sexe ll code postal FMPT1 RTA UxL de 2005	79,8	1 679 792
13	date de naissance ll sexe ll code postal FMPT1 RTA UDx de 2005	79,8	1 680 135
14	date de naissance ll sexe ll code postal FMPT1 xTA UDL de 2005	79,9	1 681 110
15	date de naissance ll sexe ll code postal FMPT1 RxA UDL de 2005	79,9	1 681 220
16	date de naissance ll sexe ll code postal FMPT1 RTx UDL de 2005	79,8	1 681 572
17	date de naissance ll sexe ll code postal FMPT1 RTA xDL de 2006	79,8	1 682 286
18	date de naissance ll sexe ll code postal FMPT1 RTA UxL de 2006	79,8	1 683 627
19	date de naissance ll sexe ll code postal FMPT1 RTA UDx de 2006	80,0	1 683 846
20	date de naissance ll sexe ll code postal FMPT1 xTA UDL de 2006	80,0	1 684 437
21	date de naissance ll sexe ll code postal FMPT1 RxA UDL de 2006	80,0	1 684 534
22	date de naissance ll sexe ll code postal FMPT1 RTx UDL de 2006	80,0	1 684 744
23	date de naissance ll sexe ll code postal FMPT1 RTA xDL de 2007	80,0	1 685 177
24	date de naissance ll sexe ll code postal FMPT1 RTA UxL de 2007	80,1	1 686 116
25	date de naissance ll sexe ll code postal FMPT1 RTA UDx de 2007	80,1	1 686 244
26	date de naissance ll sexe ll code postal FMPT1 xTA UDL de 2007	80,1	1 686 711
27	date de naissance ll sexe ll code postal FMPT1 RxA UDL de 2007	80,1	1 686 767
28	date de naissance ll sexe ll code postal FMPT1 RTx UDL de 2007	80,1	1 686 973

Nota : Les codes postaux comportent six caractères. Les trois premiers (RTA) désignent la région de tri d'acheminement; les trois derniers (UDL) désignent l'unité de distribution locale. Un « x » dans un code postal (RTA UDL) indique que le caractère a été supprimé de la clé de couplage.

Évaluation de la couverture

Les taux de couverture pour les hospitalisations toutes causes confondues dans les deux provinces étaient comparables pour les couplages *fondés sur un registre* et ceux *non fondés sur un registre*. Peu importe l'approche, les taux de couverture non pondérés étaient de 23 % au Manitoba et de 17 % en Ontario (tableau 5). Les taux de couverture pondérés étaient similaires eux aussi : au Manitoba et en Ontario, 84 % pour l'approche *fondée sur un registre*, et 82 % pour celle *non fondée sur un registre*.

Les taux de couverture variaient selon l'âge. Les taux non pondérés pour les Manitobains de 75 ans et plus étaient inférieurs de 5 ou 6 points de pourcentage au total tous âges confondus. Dans le cas des enfants du Manitoba de 1 à 4 ans, les taux pondérés étaient inférieurs au total tous âges confondus de 18 (*registre*) et 16 (*pas de registre*) points de pourcentage. Tant au Manitoba qu'en Ontario, et selon les deux approches, les taux pondérés pour les 20 à 24 ans étaient plus faibles (de 6 à 7 points de pourcentage) que le total tous âges confondus.

Tableau 3
Nombre et pourcentage de répondants ayant rempli le questionnaire détaillé du Recensement de 2006 qui ont reçu leur congé d'un hôpital de soins de courte durée, selon les approches avec registre et sans registre pour le couplage, selon certaines caractéristiques sociodémographiques, Manitoba et Ontario, 2006-2007

	Répondants au recensement ayant reçu au moins un congé d'hôpital							
	Manitoba				Ontario			
	Avec registre		Sans registre		Avec registre		Sans registre	
	en milliers	%	en milliers	%	en milliers	%	en milliers	%
Total	**18,4**	**7**	**17,9**	**6**	**118,8**	**5**	**115,7**	**5**
Sexe								
Hommes	6,8	5	6,6	5	46,6	4	45,1	4
Femmes	11,6	8	11,3	8	72,1	6	70,5	6
Groupe d'âge								
Moins d'un an	0,6	15	0,7	16	4,0	15	3,9	15
1 à 4 ans	0,4	3	0,4	3	2,6	2	2,4	2
5 à 9 ans	0,3	2	0,3	2	1,7	1	1,6	1
10 à 14 ans	0,4	2	0,4	2	1,9	1	1,8	1
15 à 19 ans	1,2	6	1,2	5	3,5	2	3,3	2
20 à 24 ans	1,4	7	1,4	7	5,3	3	5,1	3
25 à 34 ans	3,0	9	2,9	8	18,7	6	18,7	6
35 à 44 ans	2,0	5	1,9	5	13,1	3	13,2	3
45 à 54 ans	2,0	5	1,9	5	12,6	3	12,4	3
55 à 64 ans	2,0	7	1,9	7	15,3	6	15,0	6
65 à 74 ans	2,0	12	1,9	11	16,8	10	16,5	10
75 ans et plus	3,1	20	2,9	19	23,3	17	21,7	16
Quintile de revenu de la famille économique								
1 (le plus faible)	4,5	8	4,4	8	25,9	5	24,4	5
2	4,6	8	4,4	8	26,3	5	25,5	5
3	3,6	6	3,5	6	23,5	5	23,1	5
4	3,1	6	3,1	5	22,0	5	21,7	5
5 (le plus élevé)	2,6	5	2,6	5	21,0	4	20,9	4
Scolarité (18 ans et plus)								
Diplôme d'études secondaires ou un niveau supérieur	8,6	6	8,5	6	76,7	5	75,7	5
Pas de diplôme d'études secondaires	7,4	11	7,0	11	29,8	9	28,4	8
Situation d'emploi								
Occupé(e)	6,2	5	6,2	5	46,1	4	46,2	4
En chômage	0,9	8	0,8	8	3,5	4	3,3	4
Inactif(ve)	11,4	15	10,9	14	69,2	11	66,1	10
Connaissance des langues officielles								
Anglais seulement	17,0	7	16,5	7	102,7	5	100,1	5
Français seulement	0,0	7	0,0	6	0,6	6	0,6	6
Anglais et français	1,1	5	1,0	5	11,9	4	11,7	4
Ni l'anglais ni le français	0,3	8	0,3	8	3,5	7	3,3	6
Identité autochtone								
Toute identité autochtone	7,0	9	6,8	8	5,8	7	5,4	6
Indien de l'Amérique du Nord seulement	6,0	9	5,8	9	4,8	7	4,5	6
Métis seulement	1,0	6	0,9	6	0,8	6	0,8	5
Inuit seulement	0,0	7	0,0	7	0,0	4	0,0	5
Identités autochtones multiples ou autres	0,1	7	0,0	7	0,1	6	0,1	6
Non-Autochtone	11,4	6	11,1	6	113,0	5	110,3	5
Statut d'immigrant								
Immigrant	1,8	6	1,7	6	34,2	5	33,1	5
Non-immigrant	16,6	7	16,1	6	84,1	5	82,1	5
Résident non permanent	0,0	2	0,0	2	0,5	2	0,5	2
Mobilité résidentielle sur un an								
Même adresse	15,7	7	15,3	6	99,9	5	97,8	5
Adresse différente au Canada	2,0	6	1,9	6	13,9	5	13,1	5
Déménagement depuis l'extérieur du Canada	0,1	3	0,1	3	0,9	3	0,9	3
Résidence en région rurale/urbaine								
Rurale et agricole	9,4	8	9,3	8	21,7	5	22,0	6
Petit centre de population (1 000 à 29 999 habitants)	2,0	8	2,0	8	13,8	6	13,5	6
Centre de population moyen (30 000 à 99 999 habitants)	0,5	6	0,5	6	11,1	6	10,6	6
Grand centre de population (100 000 habitants et plus)	6,5	5	6,1	5	72,2	5	69,6	4

Notes : % de personnes hospitalisées = (nombre de répondants au questionnaire détaillé du recensement ayant été hospitalisés/nombre de répondants au questionnaire détaillé du recensement) *100.
Sources : Recensement de la population de 2006; Couplage recensement–Base de données sur les congés des patients (BDCP) de 2006-2007, fondé sur un registre et non fondé sur un registre.

Tableau 4
Sensibilité (vrais positifs) et spécificité (vrais négatifs) des résultats, couplage non fondé sur un registre par rapport à fondé sur un registre, Manitoba et Ontario, 2006-2007

Manitoba

Non fondé sur un registre	Fondé sur un registre (norme de référence) Couplé à la BDCP: Oui	Non	Total
Couplé à la BDCP	a = vrai positif	b = faux positif	
Oui	28 428	2 848	31 276
Non	c = faux négatif 3 898	d = vrai négatif 243 763	247 661
Total	32 326	246 611	278 937

Sensibilité a/(a+c) = 87,9 % (% de vrais positifs)
Spécificité d/(b+d) = 98,8 % (% de vrais négatifs)

Ontario

Non fondé sur un registre	Fondé sur un registre (norme de référence) Couplé à la BDCP: Oui	Non	Total
Couplé à la BDCP	a = vrai positif	b = faux positif	
Oui	106 968	9 366	116 334
Non	c = faux négatif 12 712	d = vrai négatif 2 258 865	2 271 577
Total	119 680	2 268 231	2 387 911

Sensibilité a/(a+c) = 89,4 % (% de vrais positifs)
Spécificité d/(b+d) = 99,6 % (% de vrais négatifs)

Sources : Couplages recensement–Base de données sur les congés des patients (BDCP) de 2006-2007 fondé et non fondé sur un registre.

Dans le cas des hospitalisations pour une cause précise, les taux de couverture non pondérés étaient plus près des cibles pour l'Ontario et le Manitoba que les taux de couverture pondérés. En outre, les taux de couverture non pondérés pour une cause précise avaient tendance à être plus similaires d'une approche à l'autre que ceux calculés avec des facteurs de pondération.

Discussion

Selon l'approche *fondée sur un registre*, on a couplé aux registres provinciaux de l'assurance-maladie environ 90 % des répondants au questionnaire détaillé du recensement, ce qui a permis le couplage subséquent avec la BDCP à partir des NAM. Ce taux est élevé étant donné que les études canadiennes à grande échelle ont fixé à 75 % le seuil à partir duquel les données couplées sont considérées acceptables pour la recherche[8,15,26]. À l'instar d'autres études, la présente analyse montre que les couplages axés sur le nom (approche *fondée sur un registre*) produisent des taux de couplage légèrement plus élevés que les couplages non axés sur le nom (approche *non fondée sur un registre*), mais que les deux conviennent pour la recherche[1].

Plusieurs caractéristiques associées aux taux de couplage plus faibles se dégageant de la présente analyse ont été signalées auparavant[5,42-44]. Les taux de couplage ont été relativement faibles pour les personnes ayant un faible statut socioéconomique, celles s'identifiant comme Autochtones, celles ne connaissant pas les langues officielles du Canada, les résidents des régions rurales et les personnes s'étant installées récemment au Canada.

Les résultats des couplages avec la BDCP *fondé* et *non fondé sur un registre* étaient similaires : 5 % des répondants au recensement en Ontario et 6 % à 7 % de ces personnes au Manitoba ont été couplés aux données tirées des dossiers d'hôpital. Par ailleurs, les données couplées allaient dans le sens des tendances prévues de l'utilisation des services hospitaliers du fait que des pourcentages plus élevés de pauvres, de personnes âgées et de personnes s'identifiant comme Autochtones avaient été hospitalisés[45,46]. Cela donne à penser que l'approche *non fondée sur un registre* peut produire un ensemble de données dont le niveau de qualité convient pour la recherche.

L'évaluation de la couverture a révélé une uniformité entre les approches utilisées pour le couplage. Les taux de couverture non pondérés étaient plus élevés au Manitoba qu'en Ontario, ce qui reflète le pourcentage plus élevé de Manitobains qui ont rempli le questionnaire détaillé du recensement, ainsi que des taux d'hospitalisation plus élevés au Manitoba[47]. Après l'application des poids, les taux de couverture pour le Manitoba et l'Ontario souvent étaient plus similaires.

Ce que l'on sait déjà sur le sujet

- Le couplage d'enregistrements est une façon efficace d'enrichir les ensembles de données administratives et d'enquêtes.
- La combinaison de sources de données complémentaires offre de nouvelles possibilités d'analyse.
- L'utilisation des registres provinciaux de l'assurance-maladie aux fins de couplages d'enregistrements est bien établie.
- La question de savoir si les couplages fondés sur la date de naissance, le sexe et le code postal produisent des données dont la qualité suffit pour la recherche est toutefois moins claire.

Ce qu'apporte l'étude

- La présente étude compare deux approches, l'une fondée sur un registre et l'autre non fondée sur un registre, pour le couplage de données de recensement et de données hospitalières tirées des enregistrements de la Base de données sur les congés des patients.
- Les deux approches ont produit des résultats similaires.
- Les résultats portent à croire que l'approche non fondée sur un registre constitue une option valable faute d'accès aux dossiers des registres provinciaux de l'assurance-maladie.

Tableau 5
Taux de couverture pondérés et non pondérés pour les congés d'hôpital en soins de courte durée, hospitalisations toutes causes confondues et imputables à certains diagnostics principaux, selon l'approche utilisée pour le couplage, Manitoba et Ontario, 1er avril 2006 au 31 mars 2007 inclusivement

	Manitoba					Ontario				
	Nombre de congés (estimations non couplées)	Non pondérés		Pondérés		Nombre de congés (estimations non couplées) (unlinked)	Non pondérés		Pondérés	
		Registre	Pas de registre	Registre	Pas de registre		Registre	Pas de registre	Registre	Pas de registre
	en milliers	%		%		en milliers	%		%	
Toutes causes confondues	113,8	23	23	84	82	917,7	17	17	84	82
Sexe										
Hommes	45,7	22	22	81	80	381,8	17	17	84	83
Femmes	68,1	24	24	86	83	535,8	17	17	84	83
Groupe d'âge										
Moins d'un an	2,9	23	26	75	77	25,7	16	16	81	78
1 à 4 ans	2,0	28	28	65	66	20,1	18	17	86	82
5 à 9 ans	1,4	29	29	79	79	11,6	18	17	86	84
10 à 14 ans	1,7	29	28	80	78	12,3	18	17	88	85
15 à 19 ans	4,7	34	32	87	82	24,7	17	16	79	77
20 à 24 ans	7,2	27	27	77	75	37,9	16	15	77	75
25 à 34 ans	15,9	24	24	80	79	127,4	16	16	82	82
35 à 44 ans	10,6	26	25	86	86	94,9	17	17	83	84
45 à 54 ans	11,7	24	24	85	83	95,6	17	17	84	85
55 à 64 ans	13,7	22	22	83	83	118,5	18	18	87	86
65 à 74 ans	14,7	22	22	89	87	135,9	18	17	88	86
75 ans et plus	27,3	18	17	86	82	212,1	16	16	83	78
Maladies du système circulatoire (45 ans et plus)	12,3	22	21	90	88	121,1	17	17	86	84
Sexe										
Hommes	7,0	21	21	89	88	71,0	17	17	87	86
Femmes	5,3	22	22	92	88	50,1	17	16	84	81
Groupe d'âge										
45 à 54 ans	1,3	24	24	80	82	14,1	17	17	82	85
55 à 64 ans	2,3	25	24	91	89	24,1	18	18	89	89
65 à 74 ans	2,9	23	23	92	93	30,5	18	18	90	89
75 ans et plus	5,8	19	18	90	86	52,4	17	16	83	79
Lésions traumatiques et empoisonnements	9,3	24	23	79	76	75,1	16	16	81	78
Sexe										
Hommes	5,0	24	23	76	73	37,9	16	16	80	78
Femmes	4,4	24	23	82	80	37,2	17	16	82	78
Groupe d'âge										
0 à 14 ans	0,8	31	31	76	80	5,8	18	17	84	82
15 à 24 ans	1,3	30	30	82	77	7,6	16	16	77	76
25 à 44 ans	2,1	24	24	62	61	13,9	15	15	72	71
45 à 64 ans	2,2	24	23	87	83	18,7	17	17	86	86
65 ans et plus	2,9	19	18	84	82	29,2	16	15	82	77
Grossesse, accouchement et puerpéralité (15 à 44 ans)	16,7	28	27	88	85	138,6	17	17	83	83
Groupe d'âge										
15 à 24 ans	5,5	32	30	82	77	24,9	15	15	73	71
25 à 34 ans	8,9	25	25	90	88	84,6	17	17	84	85
35 à 44 ans	2,2	26	27	93	95	29,1	17	17	86	87

Taux de couverture = (Nombre total d'hospitalisations chez les répondants au questionnaire détaillé du recensement/Nombre total d'hospitalisations selon la BDCP)*100
Notes : Les estimations non couplées et couplées de la BDCP excluent les mortinaissances, les soins autres que de courte durée, les non-Canadiens, les dates de naissance après le 16 mai 2006 (jour du recensement), les cartes d'assurance-maladie non délivrées par le Manitoba ou l'Ontario (estimations non couplées seulement), et les transferts de patients depuis une résidence pour personnes âgées.
Sources : Recensement de la population de 2006; Couplage recensement–Base de données sur les congés des patients (BDCP) de 2006-2007.

Dans les données couplées, l'hospitalisation des personnes âgées, des enfants de 1 à 4 ans et des personnes de 15 à 44 ans avait tendance à être sous-estimée. Dans une certaine mesure, cela vient du fait que la population représentée dans les données provenant du questionnaire détaillé du recensement ne correspond pas exactement aux groupes de population saisis dans les données hospitalières. Par exemple, les données couplées ne tiennent pas compte de la population vivant en établissement, mais les données hospitalières en tiennent compte en partie. Les faibles taux de couverture chez les jeunes adultes peuvent dépendre du sous-dénombrement au recensement des personnes dont les modalités relatives au logement varient et (ou) d'une couverture incomplète de certains Autochtones en Ontario[48,49].

Limites

Les données couplées comportent plusieurs limites. La présente étude ne portant que sur deux provinces, on ne connaît pas très bien les possibilités de généralisation des résultats à d'autres secteurs de compétence. Des analyses préliminaires de taux de couverture ont laissé entendre qu'il est difficile d'atteindre la signification statistique lorsque les covariables sont définies de façon trop étroite. Les résultats indiquant que la couverture diminue parfois suite à l'application de poids portent à croire que l'utilisation des poids du recensement devrait être envisagée au cas par cas.

Mot de la fin

La comparaison des approches pour le couplage fournit des preuves qu'il est possible de produire des données dont la qualité suffit pour la recherche sans avoir à recourir aux registres provinciaux de l'assurance-maladie, auxquels Statistique Canada n'a pas accès la plupart du temps. Les possibilités qu'offre le fichier couplé non fondé sur un registre en ce qui a trait à la recherche sont grandes en raison de l'échantillon représentatif de la population nationale et de la puissance statistique attribuable à sa taille et à sa couverture. Quoi qu'il en soit, les utilisateurs de données couplées devraient tenir compte de l'incidence sur leurs analyses de la méthode de couplage, des taux de couplage et de couverture, et de l'exclusion de certains groupes de population. ■

Références

1. W.E. Winkler, « Record linkage », *Sample Surveys: Design, Methods and Applications*, 29A, 2009, p. 351-380.
2. L. Lix, M. Yogendran, M. Burchill *et al.*, *Defining and Validating Chronic Diseases: An Administrative Data Approach*, Winnipeg, Manitoba, Manitoba Centre for Health Policy, 2006.
3. K. Iron, B.M. Zagorski, K. Sykora *et al.*, « Living and dying in Ontario: An opportunity for improved health information », *ICES Investigative Report*, Toronto, Institute for Clinical Evaluative Sciences, 2008.
4. V.W. Leong, S. Morgan, S.T. Wong *et al.*, « Registration for public drug benefits across areas of differing ethnic composition in British Columbia, Canada », *BMC Health Services Research* (électronique), 10(171), disponible à l'adresse http://www.biomedcentral.com/1472-6963/10/171
5. R. Wilkins, M. Tjepkema, C. Mustard *et al.*, « Étude canadienne de suivi de la mortalité selon le recensement 1991 à 2011 », *Rapports sur la santé*, 19(3), 2008, p. 27-48.
6. K. Wilkins, M. Shields et M. Rotermann, « Utilisation des hôpitaux de soins de courte durée par les fumeurs – Une étude prospective », *Rapports sur la santé*, 20(4), 2009, p. 83-93.
7. C.D. Holman, A.J. Bass, D.L. Rosman *et al.*, « A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system », *Australian Health Review*, 32(4), 2008, p. 766-777.
8. E. Ng, « Effet d'immigrant en santé et taux de mortalité », *Rapports sur la santé*, 22(4), 2011, p. 25-30.
9. S.W. Hwang, R. Wilkins, M. Tjepkema *et al.*, « Mortality among residents of shelters, rooming houses, and hotels in Canada: 11-year follow-up study », *British Medical Journal*, 339, 2009, p. b4036, doi:101136/bmj.b4036.
10. M. Tjepkema et R. Wilkins, « Espérance de vie restante à l'âge de 25 ans et la probabilité de survie jusqu'à l'âge de 75 ans selon la situation socioéconomique et l'ascendance autochtone », *Rapports sur la santé*, 22(4), 2011, p. 31-36.
11. D.G. Manuel, R. Perez, C. Bennett *et al.*, *Seven More Years: The Impact of Smoking, Alcohol, Diet, Physical Activity, and Stress on Health and Life Expectancy in Ontario, An ICES/PHO Report*, Toronto, Clinical Evaluative Sciences and Public Health Ontario, 2012.
12. B. Li, H. Quan, A. Fong *et al.*, « Assessing record linkage between health care and Vital Statistics databases using deterministic methods », *BMC Health Services Research* (électronique), 6(48), 2006, disponible à l'adresse http://www.biomedcentral.com/1472-6963/6/48
13. R. Karmel, P. Anderson, D. Gibson D *et al.*, « Empirical aspects of records linkage across multiple data sets using statistical linkage keys : the experience of PIAC cohort study », *BMC Health Services Research* 10, 2010, p. 41, disponible à l'adresse http://www.biomedcentral.com/1472-6963/10/41
14. NCSIMG. *Statistical Data Linkage in Community Services Data Collection*, Canberra, Australian Institute of Health and Welfare, 2004, disponible à l'adresse http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737419651&libID=10737419650
15. R. Fransoo, P. Martens, H. Prior *et al.*, *Adult Obesity in Manitoba: Prevalence, Associations, and Outcomes*, Winnipeg, Manitoba, Manitoba Centre for Health Policy, 2011.
16. J. Bernier et K. Nobrega, Overview of record linkage. *SYMPOSIUM 99: Combining Data from Different Sources, mai 1999*. Ottawa, Ministre de l'Industrie, 2000.
17. Australian Institute of Health and Welfare, *Comparing an SLK-based and a Name-based Data Linkage Strategy: An Investigation into the PIAC Linkage.*, Data Linkage Series N°. 11 (Catalogue CSI 11), Canberra, Australian Institute of Health and Welfare, 2011.

18. J.L. Hoving, A. Del Monaco et E. MacFarlance, « Methological issues in linking study participants to Australian cancer registries using different methods: lessons from a cohort study », *Australian and New Zealand Journal of Public Health*, 29(4), 2005, p. 378-382.

19. A. Beauchamp, A.M. Tonkin, V. Sundararajan *et al.*, « Validation of de-identified record linkage to ascertain hospital admissions in a cohort study », *BMC Medical Research Methodology* 11(42), 2011, doi:10.1186/1471-2288-11-42.

20. Australian Bureau of Statistics, *Assessing the Quality of Linking School Enrolment Records to 2011 Census Data: Deterministic Linkage Methods*, Research Paper 1351.0.55.045, disponible à l'adresse http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/FAD1F2FC23A36CE4CA257C3F00180F64/$File/1351055045_dec%202013.pdf

21. L.S. Jebamani , C.A. Burchill et P.J. Martens, « Using data linkage to identify First Nations Manitobans », *Canadian Journal of Public Health* 96(1), 2005, p. S28-S32.

22. M. Rotermann, « Évaluation de la couverture des données de l'Enquête sur la santé dans les collectivitées canadiennes couplées aux dossiers de malades hospitalisés », *Rapports sur la santé*, 20(1), 2009, p. 49-57.

23. D.P. Da Silveira et E. Artmann, « Accuracy of probabilistic record linkage applied to health databases: Systematic review » *Revista Saúde Pública* 43(5), 2009, p. 875-882.

24. A. Migowski, R.B.M. Chaves, C.M. Coeli *et al.*, « Accuracy of probabilistic record linkage in the assessment of high-complexity cardiology procedures » *Revista Saúde Pública*, 45(2), 2011, p. 269-275.

25. C. Houle, J.M. Berthelot, P. David P *et al.*, « Le projet d'appariement du Recensement et des fichiers de soins de santé du Manitoba », *Federal Committee on Statistical Methodology Papers* 1997, disponible à l'adresse http://www.statcan.gc.ca/pub/11f0019m/11f0019m1996091-fra.pdf

26. J. Brennan et M.A. Lessard, *A record linkage example: Linkage of the Citizenship and Immigration (CIC) landing file to the T1FF*, Statistics Canada Methodology Seminar, 9 avril 2013.

27. Statistique Canada, *Couplages d'enregistrements approuvés*, disponible à l'adresse http://www.statcan.gc.ca/record-enregistrement/summ-somm-fra.htm

28. Statistique Canada, *Directive sur le couplage d'enregistrements*, disponible à l'adresse http://icn-rci.statcan.ca/31/31b/31b_10c025-fra.htm

29. Statistique Canada, *Recensement de la population de 2006*, disponible à l'adresse http://www12.statcan.gc.ca/census-recensement/2006/index-fra.cfm

30. St. Pierre M, Gagné W. Création d'un prototype de la base de la population canadienne. Statistics Canada Internal Working Documentation, 2012.

31. Ministry of Health and Long-term Care. *Resource Manual for Physicians*, disponible à l'adresse http://www.health.gov.on.ca/english/providers/pub/ohip/physmanual/download/section_5.pdf

32. L.L. Roos, C.A. Mustard, J.P. Nicol *et al.*, « Registries and administrative data: Organization and accuracy », *Medical Care*, 31(3), 1993, p. 201-212.

33. L.L. Roos et J.P. Nicol, « A research registry: Uses, development, and accuracy », *Journal of Clinical Epidemiology*, 52(1), 1999, p. 39-47.

34. Institut canadien d'information sur la santé, *Métadonnées de la Base de données sur les congés des patients (BDCP)*, disponible à l'adresse http://secure.cihi.ca/cihi-ext-portal/internet/fr/document/types+of+care/hospital+care/acute+care/dad_metadata

35. Institut canadien d'information sur la santé , *Étude de l'ICIS sur la qualité des données de la Base de données sur les congés des patients pour 2005-2006*, Ottawa, Ontario, Institut canadien d'information sur la santé, 2009, disponible à l'adresse http://secure.cihi.ca/free_products/DAD_DQ_study_2005_2006_August_2009_f.pdf

36. A. Reicker. *Longitudinal Health Administrative Data (LHAD) Project*. Internal linkage of the Manitoba and Ontario health population files, documentation interne non-publiée, 2012.

37. I.P. Fellegi et A. Sunter, « A theory of record linkage », *Journal of the American Statistical Association*, 64, 1969, p. 1183-1210.

38. A. Ntwari, H. St-Jean, E. Hortop *et al.*, *Longitudinal Health Administrative Data (LHAD) Project*. External linkages between the Ontario/Manitoba health insurance registries and Ontario/Manitoba subsets of the 2006 Census of Population, documentation interne non-publiée, 2012.

39. Statistique Canada, *Rapport technique du Recensement de 2006 : Échantillonnage et pondération*, Ottawa, Ministre de l'Industrie, 2009, disponible à l'adresse http://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/sw-ep/sw-ep_index-fra.cfm

40. Institut canadien d'information sur la santé, *Les soins de santé au Canada*, Ottawa, Institut canadien d'information sur la santé, 2011, disponible à l'adresse http://secure.cihi.ca/freeproducts/HCIC_2011_seniors_report_f.pdf

41. H. Trottier, L. Martel, C. Houle *et al.*, « Vieillir chez soi ou en établissement : à quoi cela tient-il ? » *Rapports sur la santé*, 11(4), 2000, p. 49-61.

42. Australian Institute of Health and Welfare, Australian Bureau of Statistics, *National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People*, Canberra, Australian Institute of Health and Welfare, 2012, disponible à l'adresse http://www.aihw.gov.au/publication-detail/?id=60129542148

43. J. Bass et C. Garfield, « Statistical linkage keys: How effective are they? », *Proceedings of the Symposium on Health Data Linkage*, 2003, disponible à l'adresse http://www.publichealth.gov.au/pdf/reports_papers/symposium_procdngs_2003/bass.pdf

44. T. Blakely, A. Woodward et C. Salmond, « Anonymous linkage of New Zealand mortality and Census data », *Australian and New Zealand Journal of Public Health*, 2491, 2000, p. 92-95.

45. C. Sanmartin, S. Khan et l'équipe de recherche de l'Initiative sur les données longitudinales administratives et sur la santé, « Hospitalisations pour des conditions propices aux soins ambulatoires (CPSA) : les facteurs qui importent », *Série de documents de travail de la recherché sur la santé*, 2011, 7, disponible à l'adresse http://www.statcan.gc.ca/pub/82-622-x/82-622-x2011007-fra.htm

46. G. Carrière, R. Garner, C. Sanmartin *et al.*, « Hospitalisations dans des hôpitaux de soins de courte durée et identité autochtone au Canada, 2001-2002 », *Série de documents de travail de la recherché sur la santé*, 2010, 5, disponible à l'adresse http://www.statcan.gc.ca/pub/82-622-x/82-622-x2010005-fra.htm

47. Institut canadien d'information sur la santé, *Points saillants des hospitalisations et des visites aux services d'urgence en 2008-2009*, Ottawa, Institut canadien d'information sur la santé, 2010, disponible à l'adresse http://secure.cihi.ca/freeproducts/dad_nacrs_highlights_2010_05__18_f.pdf

48. Statistique Canada, *Réserves indiennes et établissements indiens partiellement dénombrés*, disponible à l'adresse http://www12.statcan.gc.ca/census-recensement/2006/ref/notes/aboriginal-autochtones-fra.cfm

49. Statistique Canada, *Rapport technique du Recensement de 2006 : Couverture*, Ottawa, Ministre de l' Industrie, 2010, disponible à l'adresse http://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/coverage-couverture/cov-couv_index-fra.cfm